

Multiple Pericentromeric Duplications of the Adrenoleukodystrophy (ALD) Locus Defines a Hotspot for Recent X to autosome Transposition Events.

Evan E. Eichler¹, Marcia L. Budarf², Mariano Rocchi³, Larry L. Deaven⁴, Norman K. Doggett⁴, David L. Nelson⁵ and Harvey Mohrenweiser¹.

¹ Human Genome Center, Lawrence Livermore National Labs; ² Children's Hospital of Philadelphia, University of Pennsylvania; ³ Instituto di Genetica, Bari, Italy ⁴ Center for Genome Studies, Los Alamos National Labs; ⁵ Human Genome Center, Baylor College of Medicine.

We recently identified a 26.5 kb duplication of the creatine transporter and DXS1357E Xq28 genes which had been differentially distributed among the pericentromeres of higher primates. Southern analysis of adjacent cosmids in this region suggested the presence of a second domain of paralogy located telomeric to the DXS1357E/creatine transporter cluster near the human adrenoleukodystrophy (ALD) gene. To investigate the nature of this second duplicated region, FISH analysis using ALD cosmids and cDNA fragments was performed, identifying at least four ALD paralogs distributed among chromosomes 2, 16, 10 and 22. PCR analysis of somatic cell deletion hybrids and yeast artificial chromosome (YAC) contigs localized three of the ALD duplicons to the cytogenetic boundaries: 10p11.1/11.2; 16p11.1/11.2; and 22q11.1/11.2, indicating an unusual propensity of the ALD locus to duplicate to pericentromere boundaries. Comparative sequencing of intronic and exonic regions among the ALD paralogs has revealed virtually identical genomic structures. An overall nucleotide similarity among the duplications was found to be 94.4%, suggesting that the ALD paralogs must have arisen within recent evolutionary time (8-10 mya). Cosmid and bacterial artificial chromosome clones have been identified corresponding to the Xq28, 10p11, 16p11 and 22q11 ALD loci. Analysis of the ALD duplication boundaries indicates that the ALD paralogy domain is approximately 10 kb in size and is located 15 kb telomeric to the DXS1357E/creatine transporter duplicated region. The 5' most breakpoint (relative to ALD transcription) is remarkably conserved among the different ALD copies and occurs near a CTGGG/CAGGG-rich sequence in intron 8. Interestingly, three out of four documented intragenic deletions among adrenoleukodystrophy patients share breakpoints near this region, suggesting that the instability of this area may be relevant to the molecular etiology of this disease. In toto, our data define a ~50 kb portion of Xq28, consisting of two regions which have undergone recent and independent duplications to specific pericentromeric autosomal loci. The distribution and dispersal of these duplicons identifies an unprecedented mechanism for targeted transposition and paralogous genome evolution.

*Performed under auspices of the US DOE by LLNL; contract No. W-7405-ENG-48.